



Attorney Docket: STOB-0001

PATENT

5

Website Management

Attorney Docket: STOB-0001

10

Inventor:

15

James A. Stob
42W704 Bridle Court
St. Charles, IL 60175
Citizen of the United States

20

Attorney:

Brian L. Longest
Cohen Mohr LLP
1055 Thomas Jefferson St, NW
Suite 504
Washington, DC 20007

25

Website Management

FIELD OF THE INVENTION

5

The present invention relates to website visibility management. More particularly to submitting webpages to Internet cataloging websites and improving website visibility.

10

RELATED APPLICATION

15

This application claims the benefit of U.S. Provisional Application serial number 60/135370, filed May 21, 1999 and entitled "Website Management".

BACKGROUND OF THE INVENTION

20

The Internet, world wide web (WWW) is growing rapidly. Websites are being added to the Internet daily and at a blazing pace. Websites are also becoming larger and it is not atypical for a website to have over 100,000 webpages or more.

25

When a website is added to the Internet it has a unique address so it may be found. The unique address is both the domain name, and the corresponding IP (Internet Protocol) address. The IP address is unique
5 to the website, as is the domain name. An IP address is typically a 32-bit number that identifies a particular network on the Internet.

Sub A² 7
10 When using a web browser you may reach an Internet site by using the IP address, eg. 209.176.240.155, or you may use the corresponding domain name, eg. Positionpro.com. A URL (Uniform Resource Locator) is the address of a file accessible on the Internet. The URL contains the name of the protocol required to
15 access the resource, in the case of web pages the protocol would be the HTTP (the Hypertext Transfer Protocol) and a domain name to identify a specific computer on the Internet, along with a file or directory path if necessary.

20 Sub A³ 7
When using a URL to view the webpages at the PositionPro website, you could use the IP address as http://209.176.240.155/, or the protocol and domain name as http://www.positionpro.com. Most users find the protocol and
25 domain name easier to remember than the IP address.

Sub A47

Each webpage within a site has a unique name, for instance there may be two webpages on a website, one entitled "contact.html" and one entitled "company.html". To reach the contact webpage you would need to use the URL <http://www.positionpro.com/contact.html>, and for the company webpage <http://www.positionpro.com/company.html>. Every webpage has a unique name.

For a person to find a website they must remember the URL or else find the URL on a website, a magazine, or newspaper etc. Websites are usually found from links on other websites, and most often found from links on Internet cataloging websites. Links are URL's which a user may click with their mouse directing the user to the webpage the link points to.

Sub A57

Internet cataloging websites, search engines, include both directories and crawling search engines. Directories may only catalog the main URL for the website, eg: <http://www.positionpro.com>. Crawling search engines typically catalog a portion or the entire website, therefore multiple URLs are cataloged, eg: <http://www.positionpro.com>, <http://www.positionpro.com/company.html>, and <http://www.positionpro.com/contact.html>.

Popular directories include Yahoo, Open Directory, Snap, LookSmart. Popular crawling search engines include: Alta Vista, Excite/AOL, Inktomi, Infoseek, Lycos, and Webcrawler.

As Internet users search for websites they type in keywords, terms, phrases, etc., into an Internet cataloging website. These searches may return 1,000, 10,000, or more webpages with those phrases. More than likely only the top 10 or 25 URLs are shown to the user without having to click a link to view another webpage. These top 10, 25, and even 50 positions are well coveted. The positions of webpages differ depending on the keywords, terms, phrases, etc., that the searcher enters and are matched with the keywords, terms, and phrases found within the code of the webpages.

Some Internet cataloging websites, crawling search engines, will crawl the Internet, known as "webcrawlers", in order to find and then index the URLs and text of the webpages that were found during the crawl. Other Internet cataloging websites, and some crawling search engines, require that someone submit the URL through a form on the Internet cataloging

website. Once the website is found the website may be searched, known as "spidering", to find additional webpages.

5 Spidering is the act of finding the original URL webpage and then following each link, a URL directing a user to the associated webpage, found within the webpage. Spiders typically do not spider farther down than one or two links from the main webpage, leaving
10 many webpages uncatalogued. Spiders also typically only follow links found within the main webpage. Links that are not on the main webpage may never be spidered.

 Since websites want traffic, users to visit their
15 site, it is very important that the webpages within a site be indexed on an Internet cataloging website. Some Internet cataloging websites do not crawl or spider, and require someone to enter each individual URL for each webpage within the website. However, this
20 is not an easy task, entering each URL manually into each Internet cataloging website is time consuming and laborious. Only a few Internet cataloging websites were mentioned, however hundreds if not thousands exist.

25

Even if someone was able to manually submit each URL from a website into all the Internet cataloging websites they wished to be indexed in, the Internet cataloging websites are not perfect and may lose URLs.

5 This requires that the URLs be resubmitted, but you never know which Internet cataloging website has lost a URL, which URL was lost, and when it was lost, unless you search the Internet cataloging websites one at a time, for each and every URL.

10

Users must also submit URLs frequently, not all Internet cataloging websites catalog every URL given to them. Internet cataloging websites also typically have daily, weekly, and monthly quotas on the number of URLs that may be submitted from a given website. Therefore,
15 it may take multiple submissions before a URL is cataloged. Someone has to keep track of how many URLs were submitted to each engine, which URLs were submitted to which engine, and when each URL was
20 submitted to which engine.

Another difficult task is keeping track of the URLs. Additional webpages are created for websites constantly, so URLs may change, new URLs may be created
25 and URL's removed. This is another time consuming

task. URL's may also be dynamic. Dynamic URL's are created at the time the user clicks on a link or otherwise requests a webpage that is automatically created by a program on the website, an example is a
5 webpage tailored to the user by placing the users name within the webpage to personalize the webpage.

With all the restrictions regarding URL submissions, submitting a URL for a webpage that was
10 submitted previously and is still in the engine should not be done, and is a waste of resources, if the URL webpage content has not changed. It is very difficult for someone to determine whether the webpage has changed since the last time it was submitted.

15 It is also very important to comply with Internet cataloging website rules for submissions. If a user submits too often, follows the wrong process, or makes other mistakes which an Internet cataloging website may discourage, the user runs the risk of having their URL
20 removed, or not cataloged in the first place, or worse their domain name may be banned from ever being catalogued.

Once a URL is catalogued within an Internet cataloging website, the owner of the URLs would like to know the ranking of each URL within each cataloging website, know when each URL's ranking changes, when a URL has been removed, and otherwise track the URLs of the website.

Sub A6 7 Services exist to submit a given website URL to a number of Internet cataloging websites. However these services simply submit a URL which is provided manually by a user. A user must determine when to submit URLs and perform a submission. For websites with a large number of URLs, 1000 or more, the process of manually submitting each URL to a service for submittal is also laborious and cumbersome. Some existing services may also submit multiple URLs to a website.

The disadvantages of the current services are solved by the present invention.

20

FEATURES AND ADVANTAGES

The present invention provides multiple advantages, including but not limited to the following:

(1) Website URL's may be resubmitted, through an automated process, using user preferences such as: time for resubmittal, date of resubmittal, after checking to see if the URL is already indexed in an Internet cataloging website, after checking to see if the indexed URL has achieved an acceptable ranking, after checking to see if the indexed URL has achieved an acceptable ranking for user specified key words;

(2) webpage titles, meta-tag descriptions, and meta-tag keywords, may be viewed for all website URLs in a unique, manageable layout so the user may determine if changes to webpages need to be made before a URL is submitted;

(3) when webpages using techniques that disallow the URL to be submitted to an Internet cataloging website, the URL may be modified so as to allow submittal;

For example webpage URLs utilizing frames may be submitted, but the webpages within the frames with the content are not viewable by the Internet cataloging website. The present invention allows submittal of webpages found within frames.

Another example is the use of an image map, an image which allows a user to chose a portion of the image by clicking on it and being sent to another

webpage through the URL associated with the chosen coordinates of the image map, if references to links are not found then the spider cannot follow the links, the current invention is capable of spidering image maps to obtain URL's.

Yet another example is the passing of parameters by webpages, which Internet cataloging engines are unable to catalog. By removing the parameter passed it is possible to create a catalogable URL;

(4) the entire website, all webpages, may be spidered;

(5) all URLs from spidered webpages may be submitted, and a user may choose not to submit some or all of the webpages, the present invention may also choose not to submit some or all of the webpages based on predetermined criteria;

(6) server logs, which are flat files containing information regarding website traffic, such as who came to the site, when they came, how they got there, if they used an Internet cataloging website - which terms did they use to search and find the URL, etc., may be used to glean valuable information which may be used to create optimized webpages in an effort to achieve more relevant search results;

(7) the present invention may also limit the links submitted to a subset of all links found on the website, either specified by the user, or determined by the present invention in an effort to follow Internet
5 cataloging engines rules;

(8) the present invention spiders the website, and spiders the entire website, unless instructed otherwise;

(9) the present invention may keep track of when
10 the website webpages were last spidered;

(10) all website webpages are tracked, both internal website links and external website links;

(11) external website links may be tracked as well, and whether or not the links are valid is also
15 tracked;

(12) an Internet catalog engine spider does not spider a page, directory, or entire site, located in a robots.txt file, while the present invention may spider the entire site, including links from webpages within a
20 webpage which is within the robots.txt file, for completeness;

(13) the present invention may save each webpage that is spidered and upon future spidering the webpages will be compared to determine whether any changes have

been made, if changes have not been made then the webpage does not have to be resubmitted;

(14) depending on Internet catalog engine rules, or at a users request, a limited number of website URLs may be submitted at any one time, based on time of day, day of month, etc.;

(15) pages may also be selectively submitted to Internet catalog engines based on whether or not they have a ranking, or an acceptable ranking, within the Internet catalog engine;

(16) the present invention spider can count levels of directories to determine how deep the spider has penetrated the website;

(17) test the webpage code to check for errors before submitting the URL to an Internet catalog engine;

(18) submittal of webpage URLs from files, instead of webpage spidering, since URLs may not be linked to a main page that would be found by the Internet catalog engine's spider;

(19) URLs may be selectively submitted, based on criteria such as the newest URL links found, last submitted, first submitted, lowest Internet catalog engine rankings in general or for specific keywords;

(20) determine how high a URL for a webpage ranks based on keywords;

(21) suggest keywords to be used based on the webpage or prior search results;

5 (22) rankings and reports show progress being made, submission strategies may be revised based on the results;

(23) allowing a file of links to be read and spidered without submitting the main file containing the links, thereby keeping the master link file anonymous and unavailable to internet catalog engines;

10 (24) when searching an Internet cataloging engine for rankings of a domain name, URL's may appear for the chosen domain which have not been found by the spider, these may be URL's which are no longer active, these URL's will be noted as found and the domain checked to determine if the URL is "not found" or what the status is, also the ranking and other statistics may be kept, and

15 (25) all of the results of the above features may be reported both on-screen and off-line, to a printer, file, database, etc.

The present invention manages website visibility. In accordance with the present invention, webpage URLs within websites will be efficiently and effortlessly submitted and catalogued with Internet cataloging search engines. A variety of features are provided to create a website and webpages which may be more easily received by the Internet cataloging website. In accordance with one feature of the invention, webpage URLs may not be submitted if the maximum number of submittals have been reached. In accordance with another feature of the invention, webpage URLs may not be submitted if the webpage has not been modified since the last submittal, unless it is no longer in the search engine. Additional features are provided for managing a websites visibility.

BRIEF DESCRIPTIONS OF THE DRAWINGS

Figures 1 through 4 are block diagrams illustrating the process of the present invention.

Figure 1 is a block diagram of the present invention process.

Figure 2 is a block diagram continuation of the present invention process in Figure 1.

Figure 3 is a block diagram continuation of the present invention process in Figure 2.

5 Figure 4 is a block diagram continuation of the present invention process in Figure 3.

Figures 5 through 32 are screen shots of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Referring now to the figures, figure 1 is a block
5 diagram of the present invention process. Step 100
begins the process with an initial spidering of a
website. It is preferred to spider the website by
moving through the directories to find the webpages,
therefore the entire website will be spidered and all
10 webpages found. Webpage URLs may be created by using
the domain name and directories to create acceptable
URLs.

Spidering by pulling URLs out of the main webpage
15 will not find webpages which are not linked off of the
main webpage or a subsequent webpage. By moving
through the directories of the website every webpage
will be uncovered and an acceptable URL created. All
the webpages within the website are obtained.

20

Step 102 then checks the robots.txt file. A
robots.txt file is a universally known file used on
websites to inform spiders and others searching through
the website which webpages should not be indexed by an

internet cataloging engine. Directories are also specified.

Step 104 then checks each individual webpage found. Step 106 determines for each webpage, whether there is a "<FRAMESET>" tag found in the webpage code. A "<FRAMESET>" tag designates that the webpage has frames. Pages source for each webpage linked off of the frame webpage needs to be found, in step 108.

10

Step 110 then determines if this is the first time the webpage has been found. If this is the first time this webpage has been found then the entire webpage may be saved into an archive area in step 112. The saving off of webpages is performed so the archived webpage may be compared to currently visible webpages on the website to determine if changes have been made that would warrant another submission to an internet cataloging search engine.

20

Step 114 is reached only if the webpage has been checked before, and therefore has an archived version. The archived version of the webpage is compared with the currently visible webpages on the website to determine if changes have been made. If changes have

been made then the page is noted to be a possible resubmission. If changes have not been made then the page is noted as not having changed.

5 Step 116 then parses the webpage code to obtain common attributes: such as the page titles, metatags containing keywords, descriptions, and other common attributes. These attributes are used by Internet cataloging engines as one indicator of relevancy when
10 retrieving search results. Therefore, webmasters like to view these attributes in a manner that is easy to read and determine what is lacking and what needs to be modified, or what is working well when comparing the ranking results to the common attributes.

15 Step 118 then checks the robots.txt file to determine if the individual webpages are listed as files not to be indexed. If the individual webpages are tagged as not to be indexed then the webpage is
20 tagged so that they will not be sent to an Internet cataloging website. If the webpage is listed as not to be followed, then the webpage is tagged so it will not be indexed, but continue to follow the file anyway for additional links.

25

Step 118 then passes to continuation step 120 which continues in figure 2 as step 200. Figure 2 is a block diagram continuation of the present invention process in Figure 1.

5

Continuation step 200 passes on to step 202. Step 202 creates a file of all the webpages found on the website. Step 202 then passes to step 204. Step 204 decides whether webpages still need to be placed in the
10 file. The process then passes to step 206

Step 206 then determines if the links found are within the current website or are external. If the links are within the current website then they are
15 placed in an internal link file. If the links are external to the website then they are placed into an external link file.

Sub A7
20 Step 208 then determines if the links found in the files will be acceptable to Internet cataloging engines. An Internet cataloging engine can only accept links that will direct a user to a webpage when clicked. A link is a URL which has the address of a file accessible on the Internet. The URL contains the
25 name of the protocol required to access the resource,

Sub A¹⁷

in the case of web pages the protocol would be the HTTP
(the Hypertext Transfer Protocol) and a domain name to
identify a specific computer on the Internet, along
with a file or directory path if necessary. For
5 example, http://www.positionpro.com/price.cfm, or
http://209.176.240.155/price.cfm.

Sub A⁸⁷

If a link, the file, does not have the domain then
add the domain name and appropriate directories. The
10 domain in this illustrative example is simply
"positionpro.com". So for a file named "price.html"
within a directory named "price", the resulting URL
would be http://www.positionpro.com/price/price.html.
This URL would be acceptable to an Internet cataloging
15 website.

Step 210 then removes links, files, which would
not be valid to submit to Internet cataloging websites.
Such invalid files would be pictures, such as JPEG and
20 GIF files, and others non-webpages. Step 212 then
begins the submittal process which continues in figure
3.

Figure 3 is a block diagram continuation of step
25 212 in figure 2. Step 300 begins the submittal process

by passing the process to step 302. Step 302 determines if there are websites in the queue to be submitted to the Internet cataloging websites. If there are not websites left to be submitted then the process ends at step 304. If additional websites are left the process passes to step 306.

Step 306 retrieves the domain name of the next website to be submitted to an Internet cataloging website. Step 308 then determines if the website may be submitted. A website may not be submitted for a variety of reasons. It is possible that the particular website is not to be submitted until the next submission process, and the user of the process can determine when websites should and should not be submitted.

If the website is to be submitted, then step 308 passes the process on to step 310. If the website is not to be submitted, the process passes back to step 302 to determine if additional websites are in the queue to be submitted.

Step 310 then determines if the website is to be submitted to the first Internet cataloging website in

the list of websites. Steps 310, 314, and 318, each determine if another Internet cataloging website is to be submitted to. In each step 310, 314, and 318, if the Internet cataloging website is to be submitted to then the process passes to step 312, 316, and 320, respectively. Each step 312, 316, and 320 then pass the process to step 400 shown in figure 4 for submittal to the Internet cataloging website.

The process works down through 310, 314, and 318, and then on to step 322 to determine if all websites have been submitted to. If additional websites need to be submitted then the process passes back to step 302. If all websites had been submitted to then the process passes on to step 324 and is finished.

Figure 4 is a block diagram continuation of the present invention process in Figure 1. Step 400 begins the process. Step 402 determines if the URL is valid. Validity not only means acceptability by an Internet cataloging website, but also whether or not the URL points to an active webpage that exists and is obtainable over the Internet. If the URL is invalid then it is flagged in step 404.

If the URL is valid then step 408 determines if the Internet cataloging website is presently working or has problems. The Internet cataloging website may be pinged by sending out a test to determine if the
5 submittal of a URL will return an error or work correctly.

If the Internet cataloging website is having problems and cannot currently receive URL submissions
10 then the process passes to step 410. Step 410 immediately sends a notification via e-mail to the administrator of the present invention to inform them that submittals cannot be made for a particular Internet cataloging website and it needs to be
15 investigated. In step 414 the process stops and is passed back to the process in figure 3 for submittal to another Internet cataloging website.

If the Internet cataloging website is working fine
20 and can currently receive URL submissions then the process passes to step 412. Step 412 determines if the maximum number of URLs have been submitted. Internet cataloging websites have rules about daily, weekly, and monthly submissions and set a maximum number of URLs
25 that may be submitted for any one particular domain.

Once that number has been met the present invention ceases the submission of URLs to that particular Internet cataloging website.

5 Step 416 marks the file of URLs for the current website domain with the last URL to be submitted. The process passes to step 414 and the process is passed back to the process in figure 3 for submittal to another Internet cataloging website.

10

 If the maximum number of URLs have not been submitted then the process passes to step 418. In step 418 the URL is submitted to the Internet cataloging website. The URL is then flagged as being submitted to
15 that particular Internet cataloging website, and the time and date of the submission is recorded. The process then passes to step 420 to wait for a response from the Internet cataloging website.

20

 Step 422 then determines if the URL was received successfully. If the URL was not received successfully then step 424 sends an email to the administrator of the present invention denoting that a problem occurred. The administrator is told which URL was to be
25 submitted, which Internet cataloging website it was to

be submitted to, date of submittal, time of submittal, and error message. The URL is also flagged as not received properly.

5 The process then passes to step 426 to determine if additional URLs need to be submitted for the website. If additional URLs need to be submitted then the process passes to step 406. Step 406 then obtains the next URL for the current website and passes the
10 process on to step 402.

 If additional URLs do not need to be submitted then the process passes from step 426 to step 428 and finishes submittal to the current Internet cataloging
15 website and the current website. The process passes back to the process in figure 3.

 Figures 5 through 32 are screen shots of the present invention.

20

 Figure 5 is a screen shot of the present invention showing the number of URLs which have been submitted to Internet cataloging websites, the number of submissions to date, and the restrictions each Internet cataloging
25 website has. Restrictions are shown as the maximum

number of submissions each Internet cataloging website is able to receive per day and per week.

The screen shot shows a list of menu items down the
5 left side of the screen as follows: Home, Main, Submissions, internal URLs, Internal Errors, Frames, Doorway, Ranked URLs, Indexed Count, Excluded URLs, External Links, External Errors, Rankings, History, Titles, Description, Keywords, Lookup/Add URL, Search
10 Engines, Edit Keywords, Retrieve code. These menu items are repeated on every screen shot.

Sub A⁹ 7
15 Figure 6 is a screen shot of statistics for the current website being submitted to Internet cataloging engines, the website is shown as <http://www.tahoevacationguide.com>. Multiple statistics are shown: 395 pages were acceptable to search engines, with 4 possible errors, 83 external links found, 1 possible error, pages without titles, descriptions,
20 keywords etc., and the total number of submissions are shown.

Figure 7 is a screen shot of the individual webpages submitted to a specific Internet cataloging
25 engine, and whether or not they were accepted.

Figure 8 is a screen shot of the individual webpages submitted to a specific Internet cataloging engine, and the status of each webpage.

5

Figure 9 is a screen shot of individual webpages that had a problem and the webpage that referenced the problematic webpage.

10

Figure 10 is a screen shot showing the webpages that have frames and when they were last crawled.

15

Figure 11 is a screen shot of the doorway pages that were last crawled.

20

Figure 12 is a screen shot showing webpages that rank within an Internet cataloging engine, which Internet cataloging engine they rank in, and the phrase that the webpage was found under when doing a query within the Internet cataloging engine.

Figure 13 is a screen shot of the URLs which have been tagged as URLs which should not be submitted to

Internet cataloging engines, either by the robots.txt file or from a 'noindex' tag.

Figure 14 is a screen shot of external links found within the website being submitted to Internet cataloging engines. The external links have codes associated with them to show if the external webpage is: not validated, okay, not found, moved, or there was a connection failure.

10

Figure 15 is a screen shot of the one external webpage that showed a code which indicated a possible problem. Code number eight shows that there was an error connecting to the external webpage. A link showing the webpage that referenced the external link is also shown for debugging purposes.

15

Sub A¹⁰ 7
Figure 16 is a screen shot of which Internet cataloging engines show a webpage from the domain name being submitted within the first 10 search results, and then within the second 10 search results. The words and phrases used when searching the Internet cataloging engines are also shown. Finally, the actual webpage that was found on each Internet cataloging engine is shown.

25

Figure 17 is a screen shot of the webpages that were ranked within a given Internet cataloging engine, the date the webpage was found, and which search result
5 page the webpage was found on. Additional information about each webpage can be found by following the "Info" link shown on the right side of the screen.

Figure 18 is a screen shot of how many webpages of
10 the domain name being submitted to the Internet cataloging engines were found within the first two pages returned by the Internet cataloging engines, on specific dates.

Figure 19 is a screen shot showing the titles of
15 all the webpages within the domain that is being submitted to Internet cataloging engines. The purpose is to show the webmaster whether they have any titles at all, or whether they are writing effective titles. In
20 many cases, as this screen shot shows, the web programmer simply used the same title for multiple webpages, which does not assist a user searching for the information found on the webpage if the title does not reflect the information found on the webpage.

25

The title is shown in the title bar of the web browser and is used frequently by Internet cataloging engines to assist in finding relevant search results. The screen shot assists in showing whether or not the web programmer is effectively using webpage titles.

Figure 20 is a screen shot showing the descriptions of all the webpages within the domain that is being submitted to Internet cataloging engines. The screen shot assists in showing whether or not the web programmer is effectively using webpage descriptions.

Figure 21 is a screen shot showing the keywords of all the webpages within the domain that is being submitted to Internet cataloging engines. The screen shot assists in showing whether or not the web programmer is effectively using webpage keywords.

Figure 22 is a screen shot of a search for webpages.

Figure 23 is a screen shot showing that the user may decide how many webpages to submit to a specific Internet cataloging website per day and per week.

Figure 24 is a screen shot showing the keywords to be searched when determining whether or not webpages from the domain are found within the Internet cataloging engines.

5

Figure 25 is a screen shot of a search capability to e-mail the code of a webpage in text format.

Sub A¹¹ 7 Figure 26 is a screen shot of detailed information
10 for a specific webpage, the webpage shown is [http://www.tahoevacationguide.com/Groups/amenitiesand rates.html](http://www.tahoevacationguide.com/Groups/amenitiesandrates.html). The user may choose which Internet cataloging engines to submit the webpage to. Title, description, and keywords are shown, along with the date
15 the webpage was first found and the date the webpage was last crawled. The webpage referring this webpage is shown. Finally the time and date of each submittal to an Internet cataloging engine is shown.

20 Figure 27 is a screen shot similar to figure 26, however this screen shot shows that the webpage has been scheduled to be submitted to three Internet cataloging engines shown with the checks next to the engines name.

25 Figure 28 is a screen shot showing similar information to that in figure 26.

Figure 32 is another screen shot of administrative functions that may be performed by the programmer maintaining the present invention.